Spatial Statistical Data Fusion for Remote Sensing Applications

Hai Nguyen[1] Matthias Katzfuss[2] Noel Cressie[3]
Amy Braverman[1]

[1]Jet Propulsion Laboratory,
California Institute of Technology

[2]Mathematics Center
Universität Heidelberg

[3]Department of Statistics,
The Ohio State University

Outline

Definition

Data fusion is the process of combining information from heterogeneous sources into a single composite picture of the relevant process, such that the composite picture is generally more accurate and complete than that derived from any single source alone (Hall, 2004).

National Aeronautics and Space Administration
Jet Propulsion Laboratory
California Institute of Technology
Pasadena, California

Motivation: remote sensing data

What is the benefit of data fusion?

- Remote sensing data are often incomplete, sparse, and spatially and temporally heterogeneous. Our goal is to infer the true physical process from all available data sources.
- Data fusion capitalizes on complementary strengths of the individual datasets to minimize prediction errors.
- Correlation in *space* and *time* can be exploited for improved accuracy.

Estimating lower atmosphere CO2

- The lower atmosphere (below 500 hPa)is where CO_2 enters and exits the atmosphere. This may be a proxy for 'sources' and 'sinks'.
- No satellite instrument currently provides measurements of CO_2 globally near the Earth's surface.
- The Greenhouse gases Observing SATellite (GOSAT) provides total-column CO_2, while the Atmospheric InfraRed Sounder (AIRS) provides mid-tropospheric CO_2.
- Approximations to lower-atmospheric CO_2 may be made by deriving joint predictions of total-column CO_2 and mid-tropospheric CO_2 and taking a (weighted) difference.

Example of satellite orbits and footprints

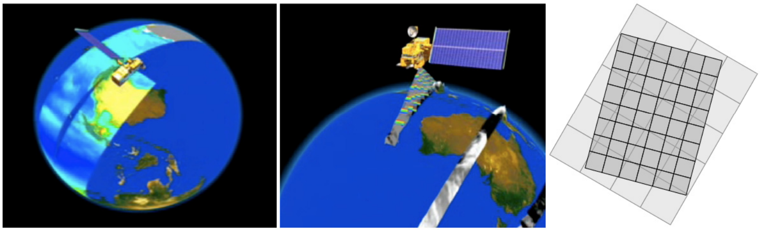


Figure: Example of different footprints (Source: Amy Braverman)

Difficulties of data fusion

Difficulties encountered when fusing remote sensing datasets:

- Massive size,
- Change of support,
- Isotropy and stationarity,
- Accounting for instruments' biases.

Outline

Introduction

Prelude: spatial (non-temporal) interpolation for massive datasets

Spatio-Temporal Data Fusion

Application to Lower Atmosphere CO2

Conclusion

Motivation

A special case of Spatio-Temporal Data Fusion is non-temporal spatial interpolation. It is convenient to introduce the ideas of STDF by talking about the spatial-only case first.

Data model

We assume the data from an instrument is generated according to the following model:

$$\begin{aligned} \mathbf{Z} &= (Z(B_1), Z(B_2), \ldots, Z(B_N))', \\ Z(B_j) &= \frac{1}{|D \cap B_j|} \left\{ \sum_{\mathbf{u} \in D \cap B_j} Y(\mathbf{u}) \right\} + \epsilon(B_j); \;\; B_j \subset \Re^d, \end{aligned}$$

where

- D is a discretized domain made up of Basic Areal Units (BAU),
- B_{ij} is the jth footprint from dataset $i (i = 1, 2)$,
- $\mathbf{Z}_i$ is the vector of response variable from dataset i,
- $Y(\cdot)$ is the true process,
- $\epsilon_i(B_{ij})$ is the error process.

Process model

We assume that the spatial process has the following linear mixed model (Cressie and Johannesson, 2008),

$$Y(\mathbf{s}) = \mathbf{t}(\mathbf{s})'\boldsymbol{\alpha} + \mathbf{S}(\mathbf{s})'\boldsymbol{\eta} + \xi(\mathbf{s}),$$

where

- $\xi(\cdot)$ is a fine-scale variation process (white noise) w/ variance σ_ξ^2,
- $\mathbf{t}(\mathbf{s})'\boldsymbol{\alpha}$ accounts for a linear trend,
- $\boldsymbol{\eta}$ is an r-dimensional Gaussian random vector $\mathrm{var}(\boldsymbol{\eta})$,
- $\mathbf{S}(\mathbf{s})$ is an r-dimensional spatial basis expansion of $\mathbf{s}$.

Covariance model

Given the linear mixed model, the covariance model is,

$$\mathbf{\Sigma} \quad \equiv \quad \mathrm{var}(\mathbf{Z}) = \mathbf{S}'\mathbf{K}\mathbf{S} + \sigma_\xi^2\mathbf{E} + \sigma_\epsilon^2\mathbf{V}.$$

where

- $\mathbf{K} = \mathrm{var}(\boldsymbol{\eta})$: fixed dimension $r \times r$,
- $\mathbf{S} \equiv (\mathbf{S}(B_1), \ldots, \mathbf{S}(B_N))'$,
- $\mathbf{E}$ and $\mathbf{V}$ are known matrices.

Prediction form

The optimal (linear unbiased) predictor of $Y(\mathbf{s})$ can be written as

$$\hat{Y}(\mathbf{s}) = \mathbf{a}'\mathbf{Z},$$

where $\mathbf{a}$ is an N-dimensional vector of kriging coefficients.

We wish to minimize,

$$\begin{aligned} E(Y(\mathbf{s}) - \hat{Y}(\mathbf{s}))^2 &= \mathrm{var}(Y(\mathbf{s}) - \mathbf{a}'\,\mathbf{Z}), \\ &= \mathrm{var}(Y(\mathbf{s})) - 2\mathbf{a}'\,\mathrm{cov}(\mathbf{Z}, Y(\mathbf{s})) + \mathbf{a}'\,\mathrm{var}(\mathbf{Z})\,\mathbf{a}, \end{aligned}$$

with respect to **a**, subject to the unbiasedness constraint,

$$\mathbf{0} = \mathbf{a}'\,\mathbf{T} - \mathbf{t}(\mathbf{s})'.$$

Solving using the method of Lagrange multipliers, the optimal kriging coefficients **a** is,

$$\mathbf{a}' = (\mathbf{c}' + (\mathbf{t}(\mathbf{s})' - \mathbf{c}'\,\boldsymbol{\Sigma}^{-1}\mathbf{T})(\mathbf{T}'\boldsymbol{\Sigma}^{-1}\mathbf{T})^{-1}\mathbf{T}')\,\boldsymbol{\Sigma}^{-1}.$$

Prediction and standard error equations

We can interpolate at a new location with the following,

$$\begin{aligned} Y(\mathbf{s})^{FRK} &= \mathbf{a}'\,\mathbf{Z}, \\ \sigma(\mathbf{s})^{SSDF} &= \left(\mathrm{E}(Y(\mathbf{s})^{SSDF} - Y(\mathbf{s}))^2\right)^{\frac{1}{2}} \\ &= \left(\mathbf{S}(\mathbf{s})'\mathbf{K}\mathbf{S}(\mathbf{s}) + \sigma_\xi^2 - 2\mathbf{a}'\left(\mathbf{S}'\mathbf{K}\mathbf{S}(\mathbf{s}) + \mathbf{b}(\mathbf{s})\right)\right. \\ &\quad \left. + \mathbf{a}'\left(\mathbf{S}'\mathbf{K}\mathbf{S} + \sigma_\xi^2\mathbf{E} + \mathbf{V}\right)\mathbf{a}\right)^{\frac{1}{2}}, \end{aligned}$$

where

$$\mathbf{b}(\mathbf{s}) = \mathrm{cov}(\boldsymbol{\xi}, \xi(\mathbf{s})).$$

This is called Fixed Rank Kriging.

Advantages of FRK

- Inversion of $\boldsymbol{\Sigma}$ is computationally scalable using the Sherman-Morrison-Woodbury formula (Hendersen and Searle, 1981),

$$\boldsymbol{\Sigma}^{-1} = \mathbf{U}^{-1} - \mathbf{U}^{-1}\mathbf{S}'\left(\mathbf{K}^{-1} + \mathbf{S}\mathbf{U}^{-1}\mathbf{S}'\right)^{-1}\mathbf{S}\mathbf{U}^{-1},$$

  where $\mathbf{U} = \sigma_{\xi}^{2}\mathbf{E} + \mathbf{V}$.
- No assumption of isotropy or stationarity.
- Handles change of support.
- Able to handle known systematic instrument biases.

Outline

Extending FRK to spatio-temporal data fusion

We develop an extension of FRK called Spatio-Temporal Data Fusion, which has the following properties,

- Ability to derive joint-estimates of two or more processes,
- Ability to exploit both spatial and temporal dependence in the data.

Main ideas behind STDF

- We account for temporal dependence using a first-oder auto-regressive model for $\boldsymbol{\eta}$.
- We do optimal predictions using a variant of the Kalman smoother.

Data model

We assume the data from an instrument is generated according to the following model:

$$Z_t^{(k)}(A) = \frac{1}{|D \cap A|} \left\{ \sum_{\mathbf{s} \in D \cap A} Y_t^{(k)}(\mathbf{s}) \right\} + \epsilon_t^{(k)}(A),$$

where

- $A \subset R^d$, $k = 1, 2$, $t = 1, 2, \ldots, T$,
- $Y_t^{(k)}(\cdot)$ is the true process,
- $\epsilon_t^{(k)}(A)$ is the error process.

Process model

We assume that the k-th process has the following form,

$$Y_t^{(k)}(\mathbf{s}) = \mathbf{x}_t^{(k)}(\mathbf{s})'\boldsymbol{\alpha}_t^{(k)} + \mathbf{S}_t^{(k)}(\mathbf{s})'\boldsymbol{\eta}_t^{(k)} + \xi_t^{(k)}(\mathbf{s}); \quad \mathbf{s} \in D.$$

where $\mathbf{x}_t^{(k)}(\cdot)$, $\boldsymbol{\alpha}_t^{(k)}$, $\mathbf{S}_t^{(k)}(\cdot)$, $\boldsymbol{\eta}_t^{(k)}$, and $\xi_t^{(k)}(\cdot)$ are defined in an analogous fashion to the corresponding spatial-only terms.

Instrument bias

To allow for bias, we assume that the measurement-error process may have non-zero mean,

$$\begin{aligned} \mathrm{E}(\epsilon_t^{(k)}(A)) &= c^{(k)}\mathrm{E}(Y^{(k)}(A)) \\ &= c^{(k)}\mathbf{x}(A)'\boldsymbol{\alpha}. \end{aligned}$$

- The multiplicative bias coefficients $\{c^{(k)} : k = 1, 2\}$ are assumed known.

Data stacking

At time t, we can stack datasets $\mathbf{Z}_t^{(1)}$ and $\mathbf{Z}_t^{(2)}$ to form a joint vector,

$$\begin{pmatrix} \mathbf{Z}_t^{(1)} \\ \mathbf{Z}_t^{(2)} \end{pmatrix} = \begin{pmatrix} \mathbf{X}_t^{(1)} & \mathbf{0} \\ \mathbf{0} & \mathbf{X}_t^{(2)} \end{pmatrix} \begin{pmatrix} \alpha_t^{(1)} \\ \alpha_t^{(2)} \end{pmatrix} + \begin{pmatrix} \mathbf{S}_t^{(1)} & \mathbf{0} \\ \mathbf{0} & \mathbf{S}_t^{(2)} \end{pmatrix} \begin{pmatrix} \eta_t^{(1)} \\ \eta_t^{(2)} \end{pmatrix} + \begin{pmatrix} \xi_t^{(1)} \\ \xi_t^{(2)} \end{pmatrix} + \begin{pmatrix} \epsilon_t^{(1)} \\ \epsilon_t^{(2)} \end{pmatrix},$$

or equivalently,

$$\mathbf{Z}_t = \mathbf{X}_t \alpha_t + \mathbf{S}_t \eta_t + \xi_t + \epsilon_t.$$

First-order auto-regressive temporal model

We assume that the covariance parameter $\boldsymbol{\eta}_t$ evolves according to a first-order auto regressive model,

$$\boldsymbol{\eta}_t | \boldsymbol{\eta}_{t-1}, \ldots, \boldsymbol{\eta}_0 \sim N_r\left(\mathbf{H}_t \boldsymbol{\eta}_{t-1}, \mathbf{U}_t\right); \; t = 1, 2, \ldots,$$

where

- The initial state is $\boldsymbol{\eta}_0 \sim N_r\left(\mathbf{0}, \mathbf{K}_0\right)$,
- The matrices $\mathbf{H}_t$ and $\mathbf{U}_t$ are called the propagator matrix and the innovation matrix.

Spatial-Temporal Data Fusion

Given data from two different instruments $\mathbf{Z}_1, \ldots, \mathbf{Z}_T$, we wish to optimally estimate the true processes at a set of locations $P \subset D$ at time $t \in \{1, \ldots, T\}$.

Assuming that the parameters $\boldsymbol{\alpha}$, $\mathbf{K}_0$, $\{(\sigma_\epsilon^{(k)})^2\}$, $\{(\sigma_\xi^{(k)})^2\}$, $\mathbf{H}_t$, and $\mathbf{U}_t$ are known, we can optimally estimate the posterior expectations and covariances for $\{\boldsymbol{\eta}_t\}$ and $\{\boldsymbol{\xi}_t^P\}$ using a variant of Kalman smoothing.

Kalman smoothing

Let $\mathbf{Z}_{1:\tilde{t}} \equiv (\mathbf{Z}'_1, \ldots, \mathbf{Z}'_{\tilde{t}})'$, we define,

- $\boldsymbol{\eta}_{t|\tilde{t}} \equiv \mathrm{E}(\boldsymbol{\eta}_t|\mathbf{Z}_{1:\tilde{t}})$,
- $\boldsymbol{\xi}^P_{t|\tilde{t}} \equiv \mathrm{E}(\boldsymbol{\xi}^P_t|\mathbf{Z}_{1:\tilde{t}})$,
- $\mathbf{P}_{t|\tilde{t}} \equiv \mathrm{var}(\boldsymbol{\eta}_t|\mathbf{Z}_{1:\tilde{t}})$,
- $\mathbf{R}^P_{t|\tilde{t}} \equiv \mathrm{var}(\boldsymbol{\xi}^P_t|\mathbf{Z}_{1:\tilde{t}})$,
- $\mathbf{W}^P_{t:\tilde{t}} \equiv \mathrm{cov}(\boldsymbol{\eta}_t, \boldsymbol{\xi}^P_t|\mathbf{Z}_{1:\tilde{t}})$.

Kalman smoothing

We first initialize $\boldsymbol{\eta}_{0|0} = \mathbf{0}$ and $\mathbf{P}_{0|0} = \mathbf{K}_0$. The one-step ahead forecasts are,

$$\begin{aligned} \boldsymbol{\eta}_{t|t-1} &= \mathbf{H}_t \boldsymbol{\eta}_{t-1|t-1} \\ \mathbf{P}_{t|t-1} &= \mathbf{H}_t \mathbf{P}_{t-1|t-1} \mathbf{H}'_{t|t}. \end{aligned}$$

The filtering quantities for $t = 1, \ldots, T$ are:

$$\begin{aligned}
\boldsymbol{\eta}_{t|t} &= \boldsymbol{\eta}_{t|t-1} + \mathbf{P}_{t|t-1}\mathbf{S}_t' \left[\mathbf{S}_t\mathbf{P}_{t|t-1}\mathbf{S}_t' + \mathbf{D}_t\right]^{-1}(\mathbf{Z}_t - \mathbf{Q}\mathbf{X}_t\boldsymbol{\alpha}_t - \mathbf{S}_t\boldsymbol{\eta}_{t|t-1}) \\
\boldsymbol{\xi}^P_{t|t} &= \mathbf{C}_t^{PZ}\mathbf{E}_t^{PZ} \left[\mathbf{S}_t\mathbf{P}_{t|t-1}\mathbf{S}_t' + \mathbf{D}_t\right]^{-1}(\mathbf{Z}_t - \mathbf{Q}\mathbf{X}_t\boldsymbol{\alpha}_t - \mathbf{S}_t\boldsymbol{\eta}_{t|t-1}) \\
\mathbf{P}_{t|t} &= \mathbf{P}_{t|t-1} - \mathbf{P}_{t|t-1}\mathbf{S}_t' \left[\mathbf{S}_t\mathbf{P}_{t|t-1}\mathbf{S}_t' + \mathbf{D}_t\right]^{-1}\mathbf{S}_t\mathbf{P}_{t|t-1} \\
\mathbf{R}^P_{t|t} &= \mathbf{C}_t^P\mathbf{E}_t^P - \mathbf{C}_t^{PZ}\mathbf{E}_t^{PZ} \left[\mathbf{S}_t\mathbf{P}_{t|t-1}\mathbf{S}_t' + \mathbf{D}_t\right]^{-1}(\mathbf{E}_t^{PZ})'(\mathbf{C}_t^{PZ})', \\
\mathbf{W}^P_{t|t} &= -\mathbf{P}_{t|t-1}\mathbf{S}_t' \left[\mathbf{S}_t\mathbf{P}_{t|t-1}\mathbf{S}_t' + \mathbf{D}_t\right]^{-1}(\mathbf{E}_t^{PZ})'(\mathbf{C}_t^{PZ})',
\end{aligned}$$

where $\mathrm{var}(\boldsymbol{\xi}_t^P) = \mathbf{C}_t^P\mathbf{E}_t^P$, $\mathrm{cov}(\boldsymbol{\xi}_t^P, \boldsymbol{\xi}_t) = \mathbf{C}_t^{PZ}\mathbf{E}_t^{PZ}$, and $\mathbf{Q}$ is a diagonal matrix with with $\{c^{(k)}\}$ along the diagonal.

Kalman smoothing

We obtain the smoothing quantities by updating "backwards" in time (i.e., for $t = T-1, T-2, \ldots, 0$):

$$\begin{aligned}
\boldsymbol{\eta}_{t|T} &= \boldsymbol{\eta}_{t|t} + \mathbf{J}_t(\boldsymbol{\eta}_{t+1|T} - \boldsymbol{\eta}_{t+1|t}) \\
\boldsymbol{\xi}^P_{t|T} &= \boldsymbol{\xi}^P_{t|t} + \mathbf{B}_t(\boldsymbol{\eta}_{t+1|T} - \boldsymbol{\eta}_{t+1|t}) \\
\mathbf{P}_{t|T} &= \mathbf{P}_{t|t} + \mathbf{J}_t(\mathbf{P}_{t+1|T} - \mathbf{P}_{t+1|t})\mathbf{J}'_t \\
\mathbf{R}^P_{t|T} &= \mathbf{R}^P_{t|t} + \mathbf{B}_t(\mathbf{P}_{t+1|T} - \mathbf{P}_{t+1|t})\mathbf{B}'_t, \\
\mathbf{W}^P_{t|T} &= \mathbf{W}^P_{t|t} + \mathbf{J}_t(\mathbf{P}_{t+1|T} - \mathbf{P}_{t+1|t})\mathbf{B}'_t,
\end{aligned}$$

where

$$\begin{aligned}
\mathbf{J}_t &\equiv \mathbf{P}_{t|t}\mathbf{H}'_{t+1}\mathbf{P}^{-1}_{t+1|t} \\
\mathbf{B}_t &\equiv -\mathbf{C}^{PZ}_t\mathbf{E}^{PZ}_t\left[\mathbf{S}_t\mathbf{P}_{t|t-1}\mathbf{S}'_t + \mathbf{D}_t\right]^{-1}\mathbf{S}_t\mathbf{P}_{t|t-1}\mathbf{H}'_{t+1}\mathbf{P}^{-1}_{t+1|t}.
\end{aligned}$$

Prediction equation

Having obtained the joint smoothing distribution of $\{\eta_t\}$ and $\{\xi_t^P\}$ given $\mathbf{Z}_1, \ldots, \mathbf{Z}_T$, the posterior mean of $\mathbf{Y}_t^P$ at the set of locations P at time t is,

$$\begin{aligned} \mathbf{Y}_{t|T}^P &= \begin{pmatrix} \mathbf{Y}_{t|T}^{(1)P} \\ \mathbf{Y}_{t|T}^{(2)P} \end{pmatrix} \\ &= \mathbf{X}_t^P \alpha_t + \mathbf{S}_t^P \eta_{t|T} + \xi_{t|T}^P. \end{aligned}$$

Prediction standard error matrix

The mean squared prediction error matrix (equivalently the posterior covariance matrix) can be calculated as:

$$\begin{aligned}
\mathbf{M}^{P}_{t|T} &\equiv \mathrm{E}\left(\left[\mathbf{Y}^{P}_{t} - \mathbf{Y}^{P}_{t|T}\right]\left[\mathbf{Y}^{P}_{t} - \mathbf{Y}^{P}_{t|T}\right]'\right) \\
&= \begin{pmatrix} \mathbf{M}^{(1,1)P}_{t|T} & \mathbf{M}^{(1,2)P}_{t|T} \\ \mathbf{M}^{(2,1)P}_{t|T} & \mathbf{M}^{(2,2)P}_{t|T} \end{pmatrix} \\
&= \mathbf{S}^{P}_{t}\mathbf{P}_{t|T}\mathbf{S}^{P'}_{t} + \mathbf{R}^{P}_{t|T} + 2\mathbf{S}^{P}_{t}\mathbf{W}^{P}_{t|T},
\end{aligned}$$

where $\mathbf{M}^{(k,m)P}_{t|T} \equiv \mathrm{E}\left(\left[\mathbf{Y}^{(k)P}_{t} - \mathbf{Y}^{(k)P}_{t|T}\right]\left[\mathbf{Y}^{(m)P}_{t} - \mathbf{Y}^{(m)P}_{t|T}\right]'\right)$.

Notes on STDF

STDF has attractive features that make it suitable for large remote sensing datasets,

- It is fast and scalable to large data inputs,
- It exploits the inter-process correlation for improved accuracy,
- It takes advantage of both *temporal* and *spatial* dependence in the data.

Outline

Introduction

Prelude: spatial (non-temporal) interpolation for massive datasets

Spatio-Temporal Data Fusion

Application to Lower Atmosphere CO2

Conclusion

Lower-atmosphere CO2

- Deriving global distribution of lower-atmosphere CO2 over time is important for studying 'sources' and 'sinks.'
- The Greenhouse gases Observing SATellite (GOSAT) provides total-column CO_2, while the Atmospheric InfraRed Sounder (AIRS) provides mid-tropospheric CO_2.
- We will derive joint predictions of total-column CO_2 and mid-tropospheric CO_2 and taking a (weighted) difference to obtain lower atmosphere CO2.

Application outline

- We select GOSAT and AIRS data over the continental United States between June and August of 2009.
- We make joint-prediction of total-column CO2 and mid-tropospheric CO2, and use weighted differencing to derive predictions of lower atmosphere CO2.
- Predictions of lower atmosphere CO2 will be compared to coincident aircraft data from NOAA.
- We also compare the performance of STDF with an alternative interpolation methodology (locally weighted regression).

Observing tracks and atmospheric sensitivity

Example of satellite orbits and sensitivity

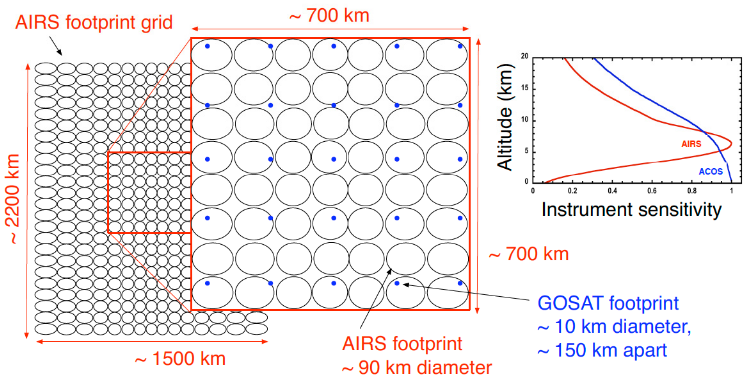


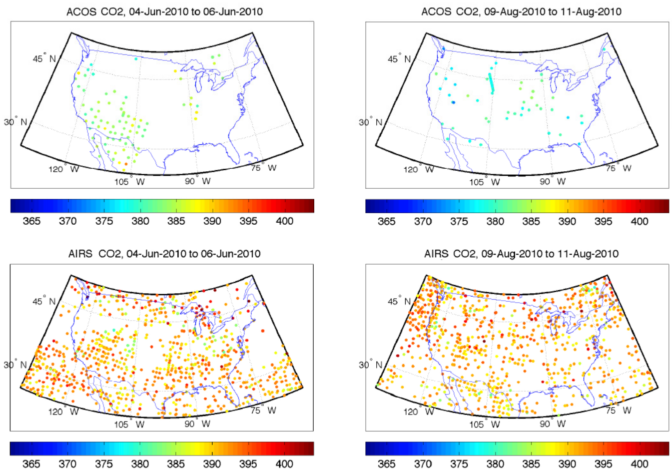
Figure: Example of GOSAT and AIRS sensitivity

ACOS and AIRS input data

- Within our domain, we have 3,869 ACOS data points and 40,564 AIRS data points.
- We group the data over these three months into 3-day blocks.
- For the elements of the vector of basis functions, we use local bisquare functions.
- The covariate function $\mathbf{t}(\cdot)$ are defined using a constant 1, latitude, and longitude.
- Given the joint prediction, $(\hat{Y}_{t|T,ACOS}(\mathbf{s}), \hat{Y}_{t|T,AIRS}(\mathbf{s}))'$, we estimate lower atmosphere CO2 as a simple linear combination

$$\hat{Y}_{t|T,LA}(\mathbf{s}) = \frac{7}{5}\hat{Y}_{t|T,ACOS}(\mathbf{s}) + \frac{2}{5}\hat{Y}_{t|T,AIRS}(\mathbf{s}).$$

Example of input data



STDF output example

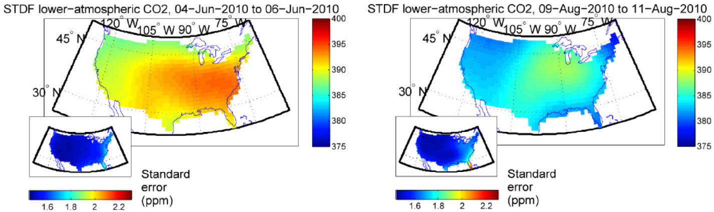


Figure: STDF output

- Run time for entire 3-month period: 4 minutes on a 3.06 GHz machine with an Intel Duo Core processor.

NOAA lower-atmosphere CO2 data

- The National Oceanic and Atmospheric Administration (NOAA) has been sampling lower-atmospheric CO_2 through a series of aircraft flights over Beaver Crossing, Nebraska and Lamont, Oklahoma.
- We can compare the NOAA aircraft data at these two locations against the corresponding 95% confidence intervals for STDF lower-atmospheric CO_2.

National Aeronautics and Space Administration

Jet Propulsion Laboratory
California Institute of Technology
Pasadena, California

NOAA lower-atmosphere CO2 comparison

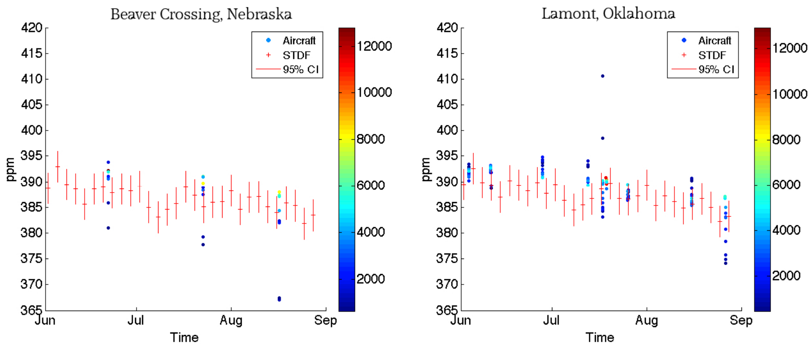
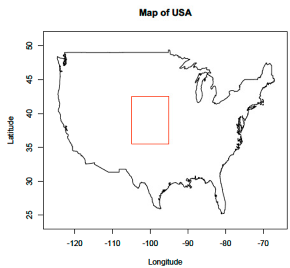


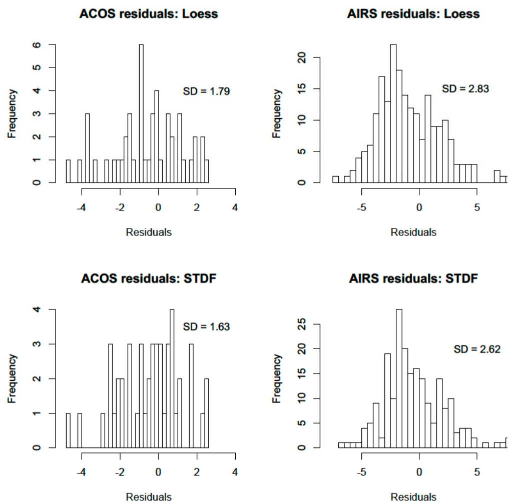
Figure: STDF outputs (red intervals) vs NOAA data (colored circles).

Comparison vs loess

- To compare against loess, we randomly 6 time blocks, and designate a small, fixed area as a reserve region.
- All data falling within the reserve region within those 6 time blocks are withheld as test data.
- We apply STDF and loess predict the value of ACOS and AIRS CO2 at the test locations.



Comparison vs loess - results



Outline

Conclusions

- STDF is fast: feasible for large remote sensing datasets.
- Results look reasonable by comparison to aircraft data for this example.
- Applicable to other types of remote sensing data, e.g., aerosols, clouds, soil moisture...
- Extensions: Bayesian inference, application to remote sensing radiances, etc.
- Questions and/or comments: contact Hai Nguyen at hai.nguyen@jpl.nasa.gov.

Acknowledgments

The authors would like to thank Colm Sweeney of the NOAA Earth System Research Laboratory for providing the in-situ aircraft CO_2 measurements used in this study.

This research is supported by NASA's Earth Science Technology Office under its Advanced Information Technology Program.

This work was partially performed at the Jet Propulsion Laboratory, California Institution of Technology under contract with the National Aeronautics and Space Administration. Government sponsorship acknowledged.

Appendix

EM estimation

Let $\boldsymbol{\theta}^{[b]}$ be the parameter vector at the b-th EM iteration. The conditional expectations and covariance matrices for the "missing data" are defined as:

$$
\begin{aligned}
\boldsymbol{\eta}_{t|T}^{[b]} &\equiv \mathrm{E}_{\boldsymbol{\theta}^{[b]}}(\boldsymbol{\eta}_t|\mathbf{Z}_{1:T}) \\
\boldsymbol{\xi}_{t|T}^{[b]} &\equiv \mathrm{E}_{\boldsymbol{\xi}^{[b]}}(\boldsymbol{\xi}_t|\mathbf{Z}_{1:T}) \\
\mathbf{P}_{t|T}^{[b]} &\equiv \mathrm{var}_{\boldsymbol{\theta}^{[b]}}(\boldsymbol{\eta}_t|\mathbf{Z}_{1:T}) \\
\mathbf{R}_{t|T}^{[b]} &\equiv \mathrm{var}_{\boldsymbol{\theta}^{[b]}}(\boldsymbol{\xi}_t|\mathbf{Z}_{1:T}) \\
\mathbf{W}_{t|T}^{[b]} &\equiv \mathrm{cov}_{\boldsymbol{\theta}^{[b]}}(\boldsymbol{\eta}_t,\boldsymbol{\xi}_t|\mathbf{Z}_{1:T}) \\
\mathbf{P}_{t,t-1|T}^{[b]} &\equiv \mathrm{cov}_{\boldsymbol{\theta}^{[b]}}(\boldsymbol{\eta}_t,\boldsymbol{\eta}_{t-1}|\mathbf{Z}_{1:T}).
\end{aligned}
$$

The cross-covariance term, $\mathbf{P}_{t,t-1|T} \equiv \mathrm{cov}(\boldsymbol{\eta}_t, \boldsymbol{\eta}_{t-1}|\mathbf{Z}_{1:T})$, is given by,

$$\begin{aligned}
\mathbf{P}_{T,T-1|T} &= (\mathbf{I}_r - \mathbf{P}_{T|T-1}\mathbf{S}'_T\left[\mathbf{S}_T\mathbf{P}_{T|T-1}\mathbf{S}'_T + \mathbf{D}_T\right]^{-1}\mathbf{S}_T) \\
&\quad \times \mathbf{H}_T\mathbf{P}_{T-1|T-1} \\
\mathbf{P}_{t,t-1|T} &= \mathbf{P}_{t|t}\mathbf{J}'_{t-1} + \mathbf{J}_t(\mathbf{P}_{t+1,t|T} - \mathbf{H}_{t+1}\mathbf{P}_{t|t})\mathbf{J}'_{t-1},
\end{aligned}$$

where $\mathbf{I}_r$ is the $r \times r$ identity matrix, and define,

$$\mathbf{L}_t^{[b+1]} \equiv \mathbf{P}_{t,t-1|T}^{[b]} + \boldsymbol{\eta}_{t|T}^{[b]}\boldsymbol{\eta}_{t-1|T}^{[b]\prime}.$$

EM estimation

The EM updates for $\boldsymbol{\theta}^{[b+1]}$ are:

$$\begin{aligned}
\alpha_t^{[b+1]} &= (\mathbf{X}_t'\mathbf{Q}\mathbf{V}_t^{-1}\mathbf{Q}\mathbf{X}_t)^{-1}\mathbf{X}_t'\mathbf{Q}\mathbf{V}_t^{-1}\left[\mathbf{Z}_t - \mathbf{S}_t\eta_{t|T}^{[b]} - \xi_{t|T}^{[b]}\right], \\
\mathbf{K}_0^{[b+1]} &= \mathbf{P}_{0|T}^{[b]} + \eta_{0|T}^{[b]}\eta_{0|T}^{[b]\prime} \\
(\sigma_{\xi,t}^{(1)})^{2\ [b+1]} &= \frac{1}{N_t^{(1)}}\text{trace}\left(\left(\mathbf{E}_t^{-1}\left[\mathbf{R}_{t|T}^{[b]} + \xi_{t|T}^{[b]}\xi_{t|T}^{[b]\prime}\right]\right)_{[1,N^{(1)}]}\right) \\
(\sigma_{\xi,t}^{(2)})^{2\ [b+1]} &= \frac{1}{N_t^{(2)}}\text{trace}\left(\left(\mathbf{E}_t^{-1}\left[\mathbf{R}_{t|T}^{[b]} + \xi_{t|T}^{[b]}\xi_{t|T}^{[b]\prime}\right]\right)_{[N^{(1)}+1,N]}\right) \\
\mathbf{H}^{[b+1]} &= \left(\sum_{t=1}^{T}\mathbf{L}_t^{[b+1]}\right)\left(\sum_{t=0}^{T-1}\mathbf{K}_t^{[b+1]}\right)^{-1} \\
\mathbf{U}^{[b+1]} &= \left(\sum_{t=1}^{T}\mathbf{K}_t^{[b+1]} - \mathbf{H}^{[b+1]}\sum_{t=1}^{T}\mathbf{L}_t^{[b+1]\prime}\right)/T.
\end{aligned}$$

EM estimation

Convergence of the parameter estimates may be monitored through the negative log-likelihood

$$-\log L(\boldsymbol{\theta}) = \frac{1}{2}\sum_{t=1}^{T} \log |\boldsymbol{\Sigma}_{\beta,t}(\boldsymbol{\theta})| + \frac{1}{2}\sum_{t=1}^{T} \boldsymbol{\beta}_t' \boldsymbol{\Sigma}_{\beta,t}(\boldsymbol{\theta})^{-1} \boldsymbol{\beta}_t .$$

Weighted differencing

We make the following simplifying assumptions,

- The air pressure at the surface of the Earth is 1000 hectopascals (hPa) and the air pressure at the satellite instrument is 0 hPa.
- The middle troposphere is the portion of the atmosphere between 500 hPa and 300 hPa.
- The CO_2 concentration above 300 hPa can be ignored.

Weighted differencing

Given total column CO_2, $Y_{\mathrm{ACOS}}(\mathbf{s})$, and mid-tropospheric CO_2, $Y_{\mathrm{AIRS}}(\mathbf{s})$, at a location $\mathbf{s}$, we approximated lower-atmospheric CO_2, $Y_{\mathrm{LA}}(\mathbf{s})$, as a simple linear combination,

$$\begin{aligned} Y_{\mathrm{LA}}(\mathbf{s}) &= \frac{(1000-300)\,Y_{\mathrm{ACOS}}(\mathbf{s}) - (500-300)\,Y_{\mathrm{AIRS}}(\mathbf{s})}{1000-500} \\ &= \frac{7}{5}Y_{\mathrm{ACOS}}(\mathbf{s}) - \frac{2}{5}Y_{\mathrm{AIRS}}(\mathbf{s}). \end{aligned}$$

Weighted differencing

From the weighted difference above, it is straightforward to obtain the prediction standard error at location **s**,

$$\sigma^2_{\mathrm{LA}}(\mathbf{s}) \equiv \left(\; 7/5, -2/5 \; \right) \mathbf{M}_{t|T}(\mathbf{s}) \left(\; 7/5, -2/5 \; \right)',$$

where $\mathbf{M}_{t|T}(\mathbf{s})$ is the prediction-error matrix for the $\mathrm{CO_2}$ prediction vector $\hat{Y}_{t|T}(\mathbf{s}) \equiv (\hat{Y}_{t|T,ACOS}(\mathbf{s}), \hat{Y}_{t|T,AIRS}(\mathbf{s}))'$.